Supplement: Non-Stochastic Bandit Slate Problems

Satyen Kale
Yahoo! Research
Santa Clara, CA
skale@yahoo-inc.com

Lev Reyzin*
Yahoo! Research
New York, NY
lreyzin@yahoo-inc.com

Robert E. Schapire†
Princeton University
Princeton, NJ
schapire@cs.princeton.edu

1 Analysis of Multiplicative Weights with Restricted Distributions

Recall our special variant of Hedge: we are allowed to uses only distributions $\mathbf{p}(t)$ from some fixed convex subset $\mathcal{P}$ of the simplex of all distributions. The goal then is to minimize regret relative to an arbitrary distribution $\mathbf{p} \in \mathcal{P}$. Such a version of Hedge is given in Figure 1, and a statement of its performance below. This algorithm is implicit in the work of [4, 6].

Algorithm $\texttt{MW}(\mathcal{P})$

Initialization: An arbitrary probability distribution $\mathbf{p}(1) \in \mathcal{P}$ on the experts, and some $\eta > 0$.
For $t = 1, 2, \ldots, T$:

1. Choose distribution $\mathbf{p}(t)$ over experts, and observe the cost vector $\boldsymbol{\ell}(t)$.
2. Compute the probability vector $\hat{\mathbf{p}}(t+1)$ using the following multiplicative update rule: for every expert i,
$$\hat{p}_i(t+1) \;=\; p_i(t)\exp(-\eta\ell_i(t))/Z(t) \tag{1}$$
where $Z(t) = \sum_i p_i(t)\exp(-\eta\ell_i(t))$ is the normalization factor.
3. Set $\mathbf{p}(t+1)$ to be the projection of $\hat{\mathbf{p}}(t+1)$ on the set $\mathcal{P}$ using the $\mathbf{RE}$ as a distance function, i.e. $\mathbf{p}(t+1) = \arg\min_{\mathbf{p}\in\mathcal{P}} \mathbf{RE}(\mathbf{p} \parallel \hat{\mathbf{p}}(t+1))$.

Figure 1: The Multiplicative Weights Algorithm with Restricted Distributions

Theorem 1.1. *Assume that $\eta > 0$ is chosen so that for all t and i, $\eta\ell_i(t) \geq -1$. Then algorithm $\texttt{MW}(\mathcal{P})$ generates distributions $\mathbf{p}(1), \ldots, \mathbf{p}(T) \in \mathcal{P}$, such that for any $\mathbf{p} \in \mathcal{P}$,*

$$\sum_{t=1}^{T} \boldsymbol{\ell}(t)\cdot\mathbf{p}(t) - \boldsymbol{\ell}(t)\cdot\mathbf{p} \;\leq\; \eta\sum_{t=1}^{T}(\boldsymbol{\ell}(t))^{\mathbf{2}}\cdot\mathbf{p}(t) + \frac{\mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(1))}{\eta}.$$

Here, $(\boldsymbol{\ell}(t))^{\mathbf{2}}$ is the vector that is the coordinate-wise square of $\boldsymbol{\ell}(t)$.

Proof. We use the relative entropy between $\mathbf{p}$ and $\mathbf{p}(t)$, $\mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(t)) := \sum_i p_i \ln(p_i/p_i(t))$ as a "potential" function. We have

$$\begin{aligned}
\mathbf{RE}(\mathbf{p} \parallel \hat{\mathbf{p}}^{t+1}) - \mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(t)) &= \sum_i p_i \ln\frac{p_i(t)}{\hat{p}_i(t+1)} \\
&= \sum_i p_i \ln\frac{Z(t)}{\exp(-\eta\ell_i(t))} \\
&= \eta(\boldsymbol{\ell}(t)\cdot\mathbf{p}) + \ln Z(t).
\end{aligned}$$

*This material is based upon work supported by the National Science Foundation under Grant #0937060 to the Computing Research Association for the Computing Innovation Fellowship program.
†This work was done while R. Schapire was visiting Yahoo! Research, New York.

Next, we have

$$\begin{aligned}
\ln Z(t) &= \ln[\sum_i p_i(t)\exp(-\eta\ell_i(t))] \\
&\leq \ln[\sum_i p_i(t)(1-\eta\ell_i(t)+\eta^2(\ell_i(t))^2)] \qquad \because \exp(-x) \leq 1-x+x^2 \text{ for } x \geq -1 \\
&= \ln[1-\eta\boldsymbol{\ell}(t)\cdot\mathbf{p}(t)+\eta^2(\boldsymbol{\ell}(t))^{\mathbf{2}}\cdot\mathbf{p}(t)] \\
&\leq -\eta\boldsymbol{\ell}(t)\cdot\mathbf{p}(t)+\eta^2(\boldsymbol{\ell}(t))^{\mathbf{2}}\cdot\mathbf{p}(t). \qquad \because \ln(1+x) \leq x \text{ for } x > -1
\end{aligned}$$

Thus, we get

$$\mathbf{RE}(\mathbf{p} \parallel \hat{\mathbf{p}}^{t+1}) - \mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(t)) \;\leq\; \eta\boldsymbol{\ell}(t)\cdot\mathbf{p} - \eta(\boldsymbol{\ell}(t)\cdot\mathbf{p}(t)) + \eta^2(\boldsymbol{\ell}(t))^{\mathbf{2}}\cdot\mathbf{p}(t).$$

Now, projection on the set $\mathcal{P}$ using the relative entropy as a distance function is a Bregman projection, and thus it satisfies the following Generalized Pythagorean inequality (see, e.g. [4]), for any $\mathbf{p} \in \mathcal{P}$:

$$\mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(t+1)) + \mathbf{RE}(\mathbf{p}(t+1) \parallel \hat{\mathbf{p}}(t+1)) \;\leq\; \mathbf{RE}(\mathbf{p} \parallel \hat{\mathbf{p}}(t+1)).$$

Since $\mathbf{RE}(\mathbf{p}(t+1) \parallel \hat{\mathbf{p}}(t+1)) \geq 0$, we get the following bound:

$$\mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(t+1)) - \mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(t)) \;\leq\; \eta\boldsymbol{\ell}(t)\cdot\mathbf{p} - \eta(\boldsymbol{\ell}(t)\cdot\mathbf{p}(t)) + \eta^2(\boldsymbol{\ell}(t))^{\mathbf{2}}\cdot\mathbf{p}(t).$$

Summing up from $t = 1$ to T, and using the fact that $\mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(T+1)) \geq 0$, and simplifying, we get

$$\sum_{t=1}^{T}\boldsymbol{\ell}(t)\cdot\mathbf{p}(t) \;\leq\; \sum_{t=1}^{T}\boldsymbol{\ell}(t)\cdot\mathbf{p} + \eta\sum_{t=1}^{T}(\boldsymbol{\ell}(t))^{\mathbf{2}}\cdot\mathbf{p}(t) + \frac{\mathbf{RE}(\mathbf{p} \parallel \mathbf{p}(1))}{\eta}.$$

□

2 The Convex Hull of Subpermutation Matrices

Recall that we represent an ordered slate π by the *subpermutation matrix* $\mathbf{M}^\pi \in \mathbb{R}^{s\times K}$ which is defined as follows: for $i = 1, 2, \ldots, s$, we have $M^\pi_{i,\pi(i)} = 1$, and all other entries are 0. In [2, 5], it is shown that the convex hull $\mathcal{M}$ of all the $\mathbf{M}^\pi$ matrices is the convex polytope defined by the following linear constraints:

$$\begin{aligned}
\forall i:\; 1 \leq i \leq s:\; \sum_{j=1}^{K} M_{ij} &= 1 \\
\forall j:\; 1 \leq j \leq K:\; \sum_{i=1}^{s} M_{ij} &\leq 1 \\
\forall ij:\; 1 \leq i \leq s, 1 \leq j \leq K:\; M_{ij} &\geq 0.
\end{aligned}$$

Clearly, all subpermutation matrices $\mathbf{M}^\pi \in \mathcal{M}$.

Given a matrix $\mathbf{M} \in \mathcal{M}$, we now wish to express it as a convex combination of $s \times K$ subpermutation matrices. For this, we complete the matrix $\mathbf{M}$ into a $K \times K$ doubly stochastic matrix $\tilde{\mathbf{M}}$ by filling out the bottom $K - s$ rows as follows: for all i, j such that $s + 1 \leq i \leq K$ and $1 \leq j \leq K$, set

$$\tilde{\mathbf{M}}_{ij} = \frac{1}{K-s}\left(1 - \sum_{i'=1}^{s} M_{i'j}\right).$$

I.e., for every column, divide up the "defect" $1 - \sum_{i'=1}^{s} M_{i'j}$ equally in the bottom $K - s$ rows. Clearly, the columns of $\tilde{\mathbf{M}}$ sum up to 1. Next, for any rows i such that $s + 1 \leq i \leq K$, we have

$$\sum_j \tilde{M}_{ij} = \sum_j \frac{1}{K-s}\left(1 - \sum_{i'=1}^{s} M_{i'j}\right) = \frac{1}{K-s}\left(K - \sum_{i'=1}^{s}\sum_{j=1}^{K} M_{i'j}\right) = 1,$$

since $\sum_{i'=1}^{s}\sum_{j=1}^{K} M_{i'j} = s$.

Now, since $\tilde{\mathbf{M}}$ is a doubly stochastic matrix, it is a convex combination of permutation matrices by Birkhoff's theorem [1]. Using the greedy algorithm described in [3] (Algorithm 1), we can decompose $\tilde{\mathbf{M}}$ into a sum of at most K^2 (actually, even at most $K^2 - 2K + 2$) permutation matrices. Now, by truncating the permutation matrices to the top s rows, we obtain (at most K^2) subpermutation matrices whose convex combination (with the same weights as in the decomposition of $\tilde{\mathbf{M}}$) is exactly $\mathbf{M}$, as required.

References

[1] BHATIA, R. *Matrix Analysis*. Springer, Berlin, 1997.

[2] BRUALDI, R. A., AND LEE, G. M. On the truncated assignment polytope. *Linear Algebra and its Applications 19* (1978), 33–62.

[3] HELMBOLD, D. P., AND WARMUTH, M. K. Learning permutations with exponential weights. In *COLT* (2007), pp. 469–483.

[4] HERBSTER, M., AND WARMUTH, M. K. Tracking the best linear predictor. *Journal of Machine Learning Research 1* (2001), 281–309.

[5] MENDELSOHN, N. S., AND DULMAGE, A. L. The convex hull of sub-permutation matrices. *Proceedings of the American Mathematical Society 9*, 2 (Apr 1958), 253–254.

[6] WARMUTH, M. K., AND KUZMIN, D. Randomized PCA algorithms with regret bounds that are logarithmic in the dimension. In *In Proc. of NIPS* (2006).